DOCUMENT RESUME

ED 296 538 EC 210 307

AUTHOR Friedman, Douglas L.; And Others
TITLE Direct Observation of Classroom Quality for Learning: Reliability and Validity.
PUB DATE Sep 87
NOTE 23p.; Paper presented at the Annual Meeting of the American Psychological Association (95th, New York, NY, August 28-September 1, 1987).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Classroom Environment; *Classroom Observation Techniques; Classroom Research; *Educational Diagnosis; Elementary Education; *Mild Disabilities; Naturalistic Observation; Student Behavior; Test Reliability; Test Validity; *Time on Task

ABSTRACT

The study evaluated the reliability and validity of SCAN, a measurement technique for assessing class involvement in the learning process (classroom climate for learning). Concurrent validity was investigated by comparing SCAN ratings to a point-sampling of on-task performance behaviors. The Engaged Time Observation Instrument (ETOI) was used to record both individual student behaviors and SCAN data. Reliability of the SCAN was determined through comparison of the recordings of two independent observers. A total of 185 observations (mean length of time 41 minutes) was made of 94 regular class students (grades 2-6) from six elementary schools in two New York City school districts. Students included 39 without handicaps and 55 with mild learning problems who were receiving resource room services. Correlations of SCAN values with actual on-task behavior of the entire sample of individual students indicated a moderate relationship (.62) between measures. Expected differences in classroom atmosphere for learning between the two school districts were also reflected in the absolute values of the SCAN means, indicating that the technique can discriminate between environmental setting variables. It is concluded that SCAN is a useful technique for conducting observational assessments of children's behavior in relation to the quality of the classroom environment for learning. (JW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED296538

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Direct Observation of Classroom Quality for Learning: Reliability and Validity

Douglas L. Friedman and Anthony A. Cancelli
Graduate School of Education
Fordham University
Roland K. Yoshida
Department of Educational and Community Programs
Queens College, City University of New York

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Douglas L. Friedman
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Running head: MEASURING CLASSROOM LEARNING QUALITY

EC 210307

Abstract

The reliability and concurrent validity of a technique used to measure the classroom atmosphere for learning was investigated. On-task behavior of 94 elementary school children with and without mild handicaps was measured by point-sampling of discrete behaviors. Concurrently, the level of entire class involvement was estimated using the SCAN technique. SCAN values were found to be highly reliable. Moderate correlations were found between the point-sampling and SCAN values. These correlations held across a variety of educational contexts. In addition, the SCAN values discriminated between contexts with varying levels of class involvement. Implications of the findings in terms of assessment are discussed.

Direct Observation of Classroom Quality for Learning: Reliability and Validity

Current practice in assessment of children with mildly handicapping conditions recognizes the need to thoroughly integrate direct observation in naturalistic settings with the more traditional approach of testing, in which a pathological problem is identified as existing within children (e.g., McKinney & Feagans, 1980; Salzinger, Antrobus, & Glick, 1980; Wiederholt, Hammill, & Brown, 1983; Ysseldyke & Christenson, 1987). This literature views deviant behavior as a complex interaction between the individual and the educational contexts in which it occurs.

The importance of the shift in the focus of assessment has been borne out in naturalistic studies that have distinguished learning environments in which children with disabilities perform as well or nearly as well as their peers without handicaps from those in which they do more poorly (Bryan, 1974; Rieth, 1983; Samuels & Miller, 1983; Krupski 1985). Such findings also appear to have more practical value in suggesting classroom interventions for remediation than procedures which identify the problem as existing within the child.

Recent studies have identified a setting variable that appears to interact significantly with the behavior of children with attentional problems, the quality of the classroom atmosphere for learning. In a quasi naturalistic study, Whalen, Henker, Collins, Finck, and Dotemoto (1979) defined the ambient atmosphere of the classroom as the level of distracting noise in the room during a lesson. They found that the task attention of nonmedicated hyperactive children was significantly less than that of medicated hyperactive children and normal comparison peers when the noise level was salient. However, noise was operationalized by playing a radio in the background during the usual lesson.

In a naturalistic study, Friedman, Cancelli, and Yoshida (1987) measured the classroom atmosphere for learning as an effect of the degree of distraction created by off-task behaviors of children in the class under typical teaching circumstances. It was found that a deficiency in attending behavior of children with learning disabilities was not a stable quality; rather, it varied significantly with the level of their class' involvement in the on-going instruction. When 75% or more of the class was engaged in the on-going instruction, children with learning disabilities were significantly more academically engaged than when less than 75% of the class was involved with the lesson. However, there were marked differences in the

degree of this variation when the data was analyzed at the level of the individual child; in fact, some children with learning disabilities were more academically engaged than the majority of their peers without handicaps when distractions were salient (Cancelli, DeLeon, and Yoshida 1987).

Research by Chow (1981), Forness and Esveldt (1975), and Walter (1983) also tends to indicate that the degree of class involvement in the lesson has a measurable influence on the academic behavior of children with handicaps. Consequently, it is important to have a means of measuring the classroom atmosphere for learning when evaluating the behavior of and planning classroom interventions for individual children with learning disabilities.

However, the use of research instruments developed for purposes of measuring classroom behavior generally requires extensive training, intensive observations, or expensive equipment (e.g., Stanley and Greenwood, 1981.) Where less complex instruments exist (e.g., Ysseldyke & Christenson, 1987), there is inadequate psychometric evidence to support their use. Hence, the practical value of existing measurement devices for school psychologists and other professionals who wish to do naturalistic evaluations appears be limited.

SCAN (Friedman, Cancelli, & Yoshida, 1987) is a measurement technique for assessing the level of class involvement in the learning process which can be related the to the behavior of individual children. SCAN ratings can also be taken across a variety of educational contexts within the classroom setting, thus permitting a finer evaluation of peer influences on the behavior of individual children. Using the SCAN technique also requires minimal training and can be employed concurrently with other classroom assessment measures.

The present study evaluated the reliability and validity of the SCAN ratings. Concurrent validity was investigated by comparing SCAN ratings to a point-sampling of behaviors of individual children's on-task performance. Point-sampling was measured with the Engaged Time Observation Instrument (ETOI) (Friedman, Cancelli, & Yoshida, 1987), based on Stanley and Greenwood's (1981) CISSAR, in which discrete behaviors were found to be validly related to student achievement.

METHOD

The data was collected as part of a larger study. A complete description of the subjects, instruments and procedure is reported in Friedman, Cancelli, and Yoshida

(1987).

Subjects

A sample of 94 students were observed in their regular classrooms as part of this study. There were 39 students without handicaps and 55 with mild learning problems. They were choosen from 29 classrooms, grades 2 through 6, in six elementary schools in two New York City public school districts. The districts differed along a number of important variables. In one district (A), all of the schools were racially isolated, had economically disadvantaged populations, and were below level in reading. All of the school sites in the study from the other district (B) had racially and economically diverse populations, and were above grade level in reading.

Regular class teachers in one of the districts (N = 15) were each asked to select three students in their class who were average learners. From among these, the observers randomly selected the subject for a particular observation. Children with handicaps represented all of the children in the six schools who were receiving resource room services, were classified as learning disabled according to federal and state regulations, which includes the criteria of average intelligence, and for whom complete data could be collected for the larger study. The mean full scale WISC-R

score for the final sample of children with handicaps was 91.2 (SD = 11.19; Range = 64 to 117).

Instrument

The Engaged Time Observation Instrument (ETOI) was used to record both individual student behaviors and SCAN data. Individual behaviors were classified as academic, management, or disruptive, and divided into thirteen all-inclusive and mutually exclusive categories. They were point-sampled every 20, 30, or 60 seconds, depending on the number of students for whom data was collected. For the purposes of analysis, appropriate academic and management behaviors were combined to get an estimate of the level of the individual child's involvement with learning process at-hand.

The SCAN was coded at every fifth recording interval throughout the observation period. Observers were trained to scan the entire class and estimate the level of total class involvement with the lesson at-hand. Involvement was operationalized as a composite of the appropriate discrete categories for academic and management behaviors as specified in the recording of individual students' behaviors. The observer's impression was coded on a 5-point scale of equal intervals, ranging from 0 to 100 percent of appropriate class involvement with the lesson.

Procedure and Amount of Data Collected

The children were observed in their regular classrooms only when their regular teacher was teaching an academic subject. The times of the observations were choosen randomly, though permission of the teacher was necessary. However, no teacher asked that an observation be rescheduled.

Each observation represented a lesson in one subject or content area and this was the unit of analysis. In all there were 185 observations with SCAN data and comparable ratings for on-task behavior of individual children. In total, there were 7,591 minutes of observation, with a mean of 41.0 minutes, and a range of from 14 to 86 minutes for each observation. There were 5,796 data points collected for individual student behaviors, with a mean of 31.33 and a range of from 10 to 96 behaviors for each observation. There were 2,832 total SCAN values recorded, with a mean 15.31 and a range of from 3 to 33 values for each observation.

The reliability of the SCAN was determined by having two independent observers, using a programmed observe-record cassette tape and dual jacked earphone, make simultaneous but independent recordings. Observers were randomly assigned to pairs for each reliability check, at random points in time throughout the data collection period.

Twenty percent (N = 38) of the total observations were checked in this way.

RESULTS

Agreement for SCAN reliability was computed using the score interval method (Hawkins & Dotson, 1975) on an interval-by-interval basis. That is, only those SCANs in which the observers recorded the same values were counted as an agreement. Reliability was then computed by dividing the number of agreements by the total number of intervals in which the SCAN was recorded and multiplying the results by 100 (Hersen & Barlow, 1976). The mean percent of agreement was 94.8, with a range of from 88.2% to 100%.

The correlation obtained between the mean SCAN values and the mean on-task rates for individual children measured concurrently with ETOI was .62 for the entire sample, indicating a moderate relationship between the measures. When compared across the separate subgroups of children and the different learning contexts, the correlations were markedly stable (Table 1). Only in the case of direct instruction for children without handicaps did the correlation vary markedly and fail to reach statistical significance. It appears that the highly restricted range of the SCAN averages as well as of the on-task percentages

for individual students in this sample reduced the upper limit of the correlation.

Insert Table 1 about here

While these figures represent a reasonable estimate of the degree of relationship between the measures, the meaning of the correlations must be referenced to the distributions of the data. The distributions for the SCAN means and means for on-task behavior for all groups of students and contexts have similar shapes; they are negatively skewed (a large frequency of scores at the top end of the distribution) and leptokurtic (the distribution is more peaked than the normal distribution). Since the distributions are similar for each of the correlations, the effect of skewness on the range of the obtained correlations was probably rather small (Carroll, 1961).

The pattern of differences in means for on-task behavior and for the SCAN values provides additional evidence for the validity of the SCAN technique (Table 2). A one-way ANOVA for on-task behavior of the three student groups was significant, F (1,184) = 22.69, MSe = 9123.290, $p < .0001$. The Tukey HSD (honestly significant different) test indicated that on-task behavior for children in one of the districts (A) was significantly different from that of

the two samples in the other district (B). A one-way ANOVA for the SCAN values related to each of the groups was also significant, F (1,184) = 61.43, MSe = 13132.400, $p < .0001$, and, as with on-task behavior, the Tukey HSD test indicated that SCAN values for district A were significantly different from those in district B. While the means for on-task behavior of the two student groups in district B were significantly different from each other, the means of the SCAN values for these two groups, though, were not, according to the Tukey HSD test. This would be expected since the SCAN values were recorded in the same classes for both groups of students.

Insert Table 2 about here

A visual inspection of the SCAN and on-task percentage means, though, indicates a constant difference between the two of approximately 10 points, with the SCAN being higher (Table 2). (Since the SCAN and on-task percentages come from different scales, a statistical test for differences was not performed.) This indicates that the SCAN probably tends to slightly overestimate the degree of student involvement in learning.

DISCUSSION

The main purpose of the present study was to evaluate the reliability and concurrent validity of the SCAN technique for assessing the classroom climate for learning. It was found that SCAN values could be recorded reliably at the criteria specified for use of a scale in screening and instructional planning (Salvia & Ysseldyke, 1985).

Correlations of the SCAN values with the actual on-task behavior of the entire sample of individual children for this sample was .62, indicating a moderate relationship between the measures. In addition, correlations were similar for observations across a number of distinct educational settings and student groupings: two school districts, which differed on a number of educationally meaningful classification factors; two distinct types of instruction, teacher directed and independent seatwork; and, two classifications of students, children with handicaps and peers without handicaps. Expected differences in the classroom atmosphere for learning between the two school districts were also reflected in the absolute values of the SCAN means, thus indicating that the SCAN technique can discriminate between environmental setting variables.

There are important practical reasons for evaluating the interaction between the child and the instructional environment, particularly when assessing the needs of and planning remediation for children with learning problems. School psychologists and other professionals need to have an empirical data base not only for understanding how a child's actual behavior may be maladaptive given a certain instructional environment, but also for making decisions about how to engineer that instructional environment for the success of the child. (Englemann, Granzin, & Severson, 1979).

However, accounting for such factors as teacher checks for understanding, opportunities for practice, nature of feedback, and clarity of directions may not be enough. There is still a superordinate context of the ambient classroom atmosphere in which learning takes place that may effectively influence both the child's perception of and response to teacher behaviors as well as the quantity and quality of the teacher behaviors themselves (Cooper & Good, 1983). When children have problems in focusing and maintaining attention to the task at hand, the need to understand the interaction of ambient distractions in the environment with the child's behavior is all the more crucial. In short, defining the broader context of the learning environment in which the behavior takes place and its relation to the behavior of other children is also a

means for understanding the behavior itself.

It may be necessary to determine, for example, if a child with a learning problem is more distractible only when a certain percent of the rest of the class is off-task. If so, the child's problem may not be so much an inherent disability as the lack of particular self-management skills. Further, the degree of distractibility may be specific to a particular type of instruction (e.g., independent seatwork) or content (e.g., mathematics). Knowledge of the exact setting would then help to form and focus the design of the intervention.

SCAN is a useful technique for conducting in vivo assessments of children's behavior in relation to the quality of the classroom environment for learning. The SCAN technique requires a minimum of training to employ and score and can easily be used in conjunction with many of the existing means for evaluating children's classroom performance. It requires little if any additional time over and above that which has already been alloted for observation. It also has the advantage of being sensitive to variety of educational contexts, thereby allowing finer evaluations of the class's influence on the individual child's behavior.

In using SCAN measurements, two factors must be kept in mind. First, it appears that the technique results in inflated scores for the class's level involvement in learning. This should be taken into account when evaluating target children's on-task behavior in relation to the entire class, so that they are not needlessly penalized if a disparity does exist. Second, for research purposes in this study, it was decided *a priori* to use a 75% cutoff level on the SCAN values when comparing the task involvement of target children to the rest of the class. This level was chosen on a rational basis, as only one other study (Walter, 1983) reported using a similar cutoff level. The practitioner needs to be sensitive to the fact that individual children have different levels of tolerance for distractions in the learning environment and that the actual level may vary from one setting to another.

References

Bryan, T.S. (1974). An observational analysis of classroom behaviors of children with learning disabilities. Journal of Learning Disabilities, 7, 35-43.

Cancelli, A. A., DeLeon, S., & Yoshida, R. K. (1987, August). Implications of context in the mainstreaming of learning disabled students. Paper presented at the convention of the American Psychological Association, New York. [First author is at Fordham University at Lincoln Center]

Chow, S.H.L. (1981). A study of academic learning time of mainstreamed learning disabled students. Final report to Bureau of Education for the Handicapped, U.S. Office of Education, Washington, DC, Grant No. G007902007. San Francisco, CA: Far West Regional Laboratory for Educational Research and Development. (ERIC Document Reproduction Service No. ED 216 467)

Cooper, H. M., & Good, T. L. (1983). Pygmalion grows up. New York: Longman.

Carroll, J. B. (1961). The nature of the data, or how to choose a correlation coefficient. Psychometrika, 26, 347-372.

Englemann, S., Granzin, A., & Severson, H. (1979). Diagnosing instruction. Journal of Special Education 13, 355-363.

Forness, S.R., & Esveldt, K.C. (1975). Classroom observation

of children with learning and behavior problems. Journal of Learning Disabilities, 8, 49-52.

Friedman, D. L., Cancelli, A. A., & Yoshida, R. K. (1987). Academic engagement of elementary school children with learning disabilities. Manuscript submitted for publication.

Hawkins, R. P., & Dotson, V. A. (1975). Reliability scores that delude: An Alice in Wonderland trip through misleading characteristics of interobserver agreement scores in interval recording. In E. Ramp & G. Semb (Eds.), Behavior analysis: Areas of research and application (pp. 359-376). Englewood Cliffs, NJ: Prentice-Hall.

Hersen, M., & Barlow, D. (1976). Single case experimental designs: Strategies for studying behavior change. New York: Pergamon Press.

Krupski, A. (1985). Variations in attention as a function of classroom task demands in learning handicapped and CA-matched nonhandicapped children. Exceptional Children, 52, 52-56.

McKinney, J.D., & Feagans, L. (1980). Learning disabilities in the classroom. Final report to Bureau of Education for the Handicapped, U.S. Office of Education, Washington, DC, Grant No. GOO-76-05224. Chapel Hill, NC: University of North Carolina at Chappel Hill, Frank Porter Graham Child Development Center.

Rieth, H.J. (1983). An analysis of the impact of

instructional time within different service delivery systems on the academic achievement of mildly handicapped children. Final report to Special Education Programs, U.S. Department of Education, Washington, DC, Grant No. G008001876. Bloomington, IN: Indiana University, Center for Innovation in Teaching the Handicapped.

Salvia, J. & Ysseldyke, J. E. (1985). Assessment in special and remedial education (3rd ed.). Boston, MA: Houghton Mifflin.

Salzinger, S., Antrobus, J., & Glick, J. (1980). The ecosystem of the "Sick" kid. In S. Salzinger, J. Antrobus, & J. Glick (Eds.), The ecosystem of the "Sick" child: Implications for classification and intervention for disturbed and mentally retarded children (pp 1 - 16). New York: Academic Press.

Samuels, S.J., & Miller, N.L. (1985). Failure to find attentional differences between learning disabled and normal children on classroom and laboratory tasks. Exceptional Children, 51, 358-375.

Stanley, S.O., & Greenwood, C.R. (1981). CISSAR: Code for instructional structure and student academic response: Observer's manual. Kansas City, KS: Junipier Gardens Children's Project, Bureau of Child Research, University of Kansas, Lawrence.

Whalen, C. K., Henker, B., Collins, B. E., Finck, D., & Dotemoto, S. (1979). A social ecology of hyperactive boys:

Medication effects in structured classroom environments. Journal of Applied Behavior Analysis, 12, 65-81.

Walter, G.A. (1983). Special education programs in District 27: Teacher activity, student activity, student time on task, and program structure. Unpublished manuscript, St. John's University, New York. (ERIC Document Reproduction Service No. ED 238 214)

Wiederholt, J.L., Hammill, D.D., & Brown, V.L. (1983). The resource teacher: A guide to effective practices (2nd ed.). Boston: Allyn and Bacon.

Ysseldyke, J. E., & Christenson, S. L. (1987). The instructional environmental scale: A comprehensive methodology for assessing an individual student's instruction. Austin, TX: Pro-Ed.

Table 1

Pearson Product-Moment Correlations Between On-Task Behavior and SCAN Means

Grouping by Student Classification and Type of Instruction	r
All students, direct instruction and seatwork	.62***
All students, direct instruction	.57***
All students, seatwork	.61***
Handicapped (district A), direct instruction and seatwork	.53***
Handicapped (district A), direct instruction	.55**
Handicapped (district A), seatwork	.45**
Handicapped (district B), direct instruction and seatwork	.52***
Handicapped (district B), direct instruction	.55*
Handicapped (district B), seatwork	.47
Nonhandicapped, direct instruction and seatwork	.45**
Nonhandicapped, direct instruction	.18
Nonhandicapped, seatwork	.77**

*p < .05. **p < .01. ***p < .001

Table 2

Combined Direct Instruction and Setawork Raw Score Means and Standard Deviations for On-Task Behavior and SCAN Values by Student Classification

On-Task Behavior								
HC (district A)			HC (district B)			NH (district B)		
Mean	SD	n	Mean	SD	n	Mean	SD	n
60.1	23.5	72	76.3	19.5	68	84.3	14.0	45
SCAN Values								
68.6	21.0	72	92.2	9.4	68	94.1	6.1	45

Note. HC = children with handicaps; NH = children without handicaps